Transplant International ISSN 0934-0874

ORIGINAL ARTICLE

Evaluating the accuracy of functional biomarkers for detecting histological changes in chronic allograft nephropathy

Serdar Yilmaz,[1,2] Ipek Isik,[1,2,3] Marjan Afrouzian,[3] Mauricio Monroy,[1,2] Aylin Sar,[1,2,3] Hallgrimur Benediktsson[3] and Kevin McLaughlin,[2,4]

1 Division of Transplantation, University of Calgary, Calgary, AB, Canada
2 Southern Alberta Transplant Program, University of Calgary, Calgary, AB, Canada
3 Department of Pathology, University of Calgary, Calgary, AB, Canada
4 Division of Nephrology, University of Calgary, Calgary, AB, Canada

Keywords
chronic allograft damage index, chronic allograft nephropathy, estimated glomerular filtration rate, protocol biopsy, receiver operator characteristic curve analysis, serum creatinine, surrogate marker.

Correspondence
Dr Serdar Yilmaz, Division of Transplantation, Department of Surgery, Foothills Medical Centre, 1403-29 St N.W., Calgary, AB, Canada, T2N 2T9. Tel.: 403 944 4266; fax: 403 944 4693; e-mail: serdar.yilmaz@calgaryhealthregion.ca

Received: 8 February 2007
Revision requested: 28 February 2007
Accepted: 10 April 2007

doi:10.1111/j.1432-2277.2007.00494.x

Summary

The most common cause of late kidney transplant failure is chronic allograft nephropathy (CAN). Much research has focused on identifying biomarkers (or correlates) that would predict subsequent CAN and allow timely intervention. Functional biomarkers such as serum creatinine and estimated glomerular filtration rate (eGFR) have been widely adopted, even though they have not been rigorously evaluated as surrogate markers. This study evaluated serum creatinine and eGFR for predicting the early histopathological changes seen in transplant protocol biopsies (TPB). We prospectively followed 289 kidney transplant patients in the Southern Alberta Transplant Program who had TPB at 6–12 months post-transplant. Tissue samples ($n = 280$) were independently examined by renal pathologists. The ability of serum creatinine or eGFR to predict the threshold level for abnormal histopathology was evaluated by calculating the area under the receiver operator characteristic curve. Serum creatinine and eGFR had poor predictive value (most confidence intervals included 0.5, indicating no predictive ability) for ten individual histological measurements (Banff 97 scores), and the Chronic Allograft Damage Index. We conclude that serum creatinine and eGFR have a limited clinical role in predicting the early histopathological changes that precede CAN and should not be used for this purpose.

Introduction

Over the last two decades, advances in immunosuppressive medications have improved short-term renal graft survival. An analysis of the United Network for Organ Sharing/Organ Procurement and Transplantation Network transplant databases showed that long-term renal graft survival also improved [1]. The most common cause of late kidney graft failure is chronic allograft nephropathy (CAN) [2], and reducing the incidence of this post-transplant complication presents the greatest challenge to those involved in the care of renal transplant patients. Strategies aimed at reducing renal allograft loss due to CAN would benefit from the availability of a biomarker (or correlate) that would be visible early in the post-transplant course and predict subsequent allograft dysfunction and allograft loss due to CAN. To qualify as a true surrogate, this biomarker should lie on the causal pathway of CAN and capture the net effect of treatment on clinical end-points (allograft dysfunction or loss due to CAN) [3–5].

Early histopathological changes detected on transplant protocol biopsies (TPBs) are structural biomarkers for

the clinical end-points of allograft dysfunction and/or failure due to CAN [6–13]. Studies suggest that these changes also capture the effect of treatment [14,15], thus fulfilling the criteria for surrogate markers. However, many centers are reluctant to perform TPB, as this is an invasive procedure. Although the risk of significant complications following renal biopsy has decreased over the years, particularly with the use of ultrasound guidance [16] and automated core biopsy systems [17], there are still potential risks [18–20]. Therefore, renal biopsy is not used consistently in routine clinical management; instead, functional biomarkers are used as possible predictors of CAN. Functional biomarkers such as serum creatinine [21–23] and estimated glomerular filtration rate (eGFR) [24] clearly have the advantage of being inexpensive and non-invasive screening tests for CAN compared with TPB. However, the aforementioned studies mainly used multivariable analyses to reach their conclusions about the usefulness of these functional markers [21–24]. Although multivariable models are appropriate for identifying prognostic factors [25], further validation is required to determine whether a biomarker represents a surrogate end-point that can be used for diagnosis and assessment of interventions [26]. In the case of serum creatinine and eGFR, this type of validation has not been reported; therefore, it is our opinion that serum creatinine and eGFR have been adopted prematurely as surrogate or predictive markers of CAN.

Sensitivity, specificity, and predictive value have long been used as indices of test accuracy [27] and to validate putative surrogate markers. Newer methods, such as receiver operator characteristic curve (ROC) analysis [28,29] and likelihood ratios [27], are more robust indicators that overcome many limitations of the traditional indices. The objective of this study was to evaluate the accuracy of serum creatinine and eGFR, measured within the first year after transplant, in predicting the early abnormal histopathological changes seen in TPBs (Banff 97 scoring) using ROC analysis. The time course of this study did not allow assessment of the clinical end-point – graph loss due to CAN – therefore TPB histopathology was used as the gold standard.

Patients and methods

Study sample

Since 1998, we have been routinely performing prospective TPBs in consenting patients to allow early diagnosis of subclinical acute and chronic histological changes and subsequent intervention. These biopsies are marked as 'protocol biopsies' (i.e. TPBs) in our database if serum creatinine is stable and less than 15% during the last 1 month prior to the prescheduled biopsy date. In contrast, biopsies performed because of proteinuria or suspected acute rejection or BK nephropathy are marked as 'diagnostic biopsies' in our database and were not used in this study. The study examined all consecutive renal transplant patients in the Southern Alberta Transplant Program from July 1998 to January 2006 who underwent TPB within 6–12 months after kidney transplant. Excluded patients were those with biopsy contraindications (those who had a transplant placed in their peritoneal cavity, were on anticoagulants, or were Jehovah Witness) and patients who refused the protocol biopsies. Of 289 patients, 54% were male, and 3% were black. The mean age at the time of TPB was 47 ± 12 years.

Of the 280 patients whose biopsies were included in this study, 67% ($n = 187$) received kidneys from a deceased donor and 33% ($n = 93$) received kidneys from a living donor. The mean age of the deceased donors at the time of organ procurement was 35 ± 18 years, and the mean body mass index was 24 ± 5. Of these deceased donors, 61% were male and 89% were Caucasian. The history of hypertension was 13% (8% unknown by the consenting relative), and the history of diabetes was 2% (24% unknown by the consenting relative). The causes of death were trauma (49%), cerebrovascular accident (34%), anoxia/hypoxia (11%) and other (6%). The mean age of the living donors at the time of organ procurement was 42 ± 12 years and the mean body mass index was 26 ± 4. Of these living donors, 43% were male and 88% were Caucasian. There was no history of hypertension, diabetes, or coronary artery and peripheral vascular disease among the living donors.

During this time period, the standard induction regime included the use of anti-interleukin-2 monoclonal antibody for patients not considered to be at high immunological risk. Polyclonal induction was used for those considered at high immunological risk or those patients who experienced delayed graft function (need for dialysis more than 24 h after the transplantation). Standard maintenance therapy comprised a triple therapy regime of calcineurin inhibitor, an anti-proliferative agent (primarily mycophenolate mofetil) and prednisone. Rejection episodes were treated with solumedrol (500 mg for 3 days followed by 250 mg for 1 day) and a tapering dose of prednisone on fifth day started at 1 mg/kg. Steroid-resistant rejection was treated with anti-thymocyte globulin with intravenous immunoglobulin with the addition of plasmapheresis if there was evidence of antibody-mediated rejection.

All subjects gave their informed consent prior to their transplantations for evaluating the quality of our programme and in developing future diagnostic modalities.

Functional biomarkers

The average serum creatinine level for each patient was calculated by averaging all serum creatinine measurements taken from that patient within 10 days of the biopsy date. An average of 2.9 measurements/patient (820 samples from 280 patients) were made over these 20 days. The Modification of Diet in Renal Disease equation was used for estimating eGFR from serum creatinine.

Biopsy scores

Transplant protocol biopsies consisted of two cores obtained with 18-gauge needles using ultrasound guidance in the Radiology Department, Foothills Hospital, Calgary, Alberta. Paraffin and plastic sections were prepared and stained with haematoxylin-eosin, trichrome, periodic acid-Schiff and periodic acid-Schiff-methanamine silver. All TPBs were independently examined by light microscopy by two pathologists (I.I., A.S.). They were blinded to therapy as well as initial diagnosis of the cases. Of the 289 TPBs done over the 8-year period, nine TPBs contained fewer than seven glomeruli and were excluded from the analysis, leaving 280 protocol biopsies with seven or more glomeruli.

The following histological parameters were graded from 0 to 3 using the Banff 97 thresholds [30]: mononuclear cell interstitial inflammation (i), allograft glomerulitis (g), intimal arteritis (v), tubulitis (t), interstitial fibrosis (ci), allograft glomerulopathy (cg), fibrous intimal thickening (cv), tubular atrophy (ct), mesangial matrix increase (mm) and arteriolar hyalinosis (ah). From the semi-quantitative scoring system of the Banff 97 schema, three different threshold values were selected to create dichotomous dependent variables: ≥1 (mild), ≥2 (moderate) and 3 (severe) changes for each of the histological parameters. As well, chronic/sclerosing allograft nephropathy, which ranks the presence of interstitial fibrosis (ci) and tubular atrophy (ct), was graded as 0 (normal), I (mild), II (moderate) and III (severe) according to the Banff 97 schema.

To score the overall severity of renal allograft damage, a Chronic Allograft Damage Index (CADI) score was calculated. This score is a composite of the major Banff scores [(i), (ci), (cv), (mm) and (ct)], but uses a score for glomerular sclerosis instead of (cg) [31,32]. When each one of these parameters is scored from 0 to 3 according to Banff 97 thresholds [30], the maximal total score is 18 and the theoretical minimum is 0. Previously, patients with a CADI score >4 have been identified as having increased risk of graft loss at 3 years [15]; therefore, this was used as the threshold for defining a level of CAN severity that is associated with short-term allograft loss.

Statistical analyses

For data collection and analysis, the Southern Alberta Transplant Program real-time Kidney/Pancreas Transplant Database (ALTRAbase) was used. Kappa values [33] were used to assess the level of inter-observer agreement on the histological scoring of 30 randomly selected samples. For other analyses, the dependent variable was the threshold level for the histological parameter of interest. The independent variable was serum creatinine or eGFR. Relationship between serum creatinine or eGFR and the CADI score were investigated using Pearson product-moment correlation coefficient. The ability of the independent variable to predict the threshold level for abnormal histopathology was estimated using area under the receiver operator characteristic curve (AUROC) analysis [28,29]. In our analysis, sensitivity versus 1-specificity (true-positive versus false-positive) pairs were plotted for every creatinine or eGFR reading using various histopathological thresholds as the gold standard, and the AUROC was calculated. A test with perfect discrimination (true-positive fraction is 1, false-positive fraction is 0) has an AUROC of 1, whereas an AUROC of 0.5 (a 45° line) represents a test with no discrimination. All statistical analyses were performed using STATA 8.0 software (Stata Corporation, College Station, TX, USA) and SPSS 14.01 software (SPSS Inc., Chicago, IL, USA).

Results

Two hundred and eighty biopsy results were included in the study. At the time of TPB, the mean serum creatinine was 122 ± 38 μmol/l and eGFR was 58 ± 19.0 ml/min. The mean time interval between transplantation and TPB was 225 ± 63 days. The prevalence and severity of the acute and chronic histological changes on TPB of these patients with stable renal allograft function are shown in Table 1. Biopsies were also graded according to the Banff 97 Chronic/Sclerosing Allograft Nephropathy categories. Inter-observer kappa values (0 = no agreement, 1 = perfect agreement) and probability of agreement were 1 ($P < 0.0001$) for cg, 0.87 ($P < 0.0001$) for cv, 0.74 ($P < 0.0001$) for ci, 0.68 ($P < 0.0001$) for mm and glomerular sclerosis, 0.65 ($P < 0.0001$) for ct, 0.53 ($P = 0.001$) for g, 0.48 ($P = 0.005$) for i, 0.42 ($P = 0.016$) for ah and 0.28 ($P = 0.13$) for t, indicating high agreement between the two pathologists for most variables. Of 280 patient with stable kidney function, 190 (68%) had Grade I or more chronic allograft nephropathy, and 90 (32%) patients had no nephropathy according to the Banff 97 classification (Table 1).

Receiver operator characteristic curve plots were generated for serum creatinine levels and eGFR using either

Table 1. Prevalence and severity of histopathologic lesions and chronic/sclerosing allograft nephropathy score of 280 transplant protocol biopsies at 6–12 months post-transplant.

	Categories based on Banff 97 scoring			
Histological parameter*	Normal	Mild (≥1)	Moderate (≥2)	Severe (3)
g	220 (79%)	45 (16%)	15 (5%)	0
i	193 (69%)	74 (26%)	10 (4%)	3 (1%)
t	218 (78%)	29 (10%)	23 (8%)	10 (4%)
v	280 (100%)	0	0	0
ah†	147 (53%)	100 (36%)	26 (9%)	6 (2%)
cg	275 (98%)	5 (2%)	0	0
ci	139 (50%)	120 (43%)	18 (6%)	3 (1%)
ct	97 (35%)	165 (59%)	15 (5%)	3 (1%)
cv†	146 (52%)	111 (40%)	22 (8%)	0
mm	85 (30%)	116 (41%)	55 (20%)	24 (9%)
	Grade 0	Grade I	Grade II	Grade III
Chronic/sclerosing allograft nephropathy score	90 (32.1%)	168 (60%)	18 (6.4%)	4 (1.4%)

*g, early type of allograft glomerulitis; i, interstitial inflammation; t, tubulitis; v, intimal arteritis; ah, arteriolar hyalinosis; cg, allograft glomerulopathy; ci, interstitial fibrosis; ct, tubular atrophy; cv, fibrous intimal proliferation; mm, glomerular mesangial matrix increase.
†One sample was excluded, as the structure to be evaluated could not be seen ($n = 279$).

'mild changes' (Banff 97 score of ≥1) or 'moderate changes' (Banff 97 score of ≥2) for each of the 10 histological measurements as the diagnostic gold standard (Table 2). Area under the receiver operator characteristic curve (AUROC) values ranged from 0.51 (representing no discrimination) to 0.62, and most of the 95% confidence intervals included or were close to 0.5, indicating that neither serum creatinine nor eGFR had predictive value for any of the individual histological parameters, with the exception of interstitial inflammation (i).

We next examined the ability of serum creatinine and eGFR to predict the CADI score. Scatter plots of serum creatinine (Fig. 1a) or eGFR (Fig. 2a) against the CADI score indicated a very weak relationship ($R = 0.21$ and -0.20 respectively). Only a 5% and 4% variability in CADI scores is explained by serum creatinine or eGFR levels ($R^2 = 0.047$ and 0.042 respectively). When we considered a CADI score of >4 as the dependent variable, the AUROC was 0.55 (95% confidence interval 0.47–0.64) for serum creatinine (Fig. 1b) and 0.56 (95% confidence interval 0.48–0.65) for eGFR (Fig. 2b), indicating that neither of these measures were useful in predicting the CADI score.

Table 2. Prevalence and severity of chronic histopathologic lesions on 280 TPB 6–12 months post-transplant and the ability of serum-creatinine and eGFR to predict these changes.

	≥Mild changes (Banff 97 score ≥1)		≥Moderate changes (Banff 97 score ≥2)	
Histological parameter*	AUROC for serum-creatinine†	AUROC for eGFR†	AUROC for serum-creatinine†	AUROC for eGFR†
g	0.60 (0.52–0.69)	0.56 (0.48–0.65)	0.52 (0.37–0.77)	0.55 (0.41–0.68)
i	0.58 (0.51–0.65)	0.61 (0.53–0.68)	0.75 (0.63–0.87)	0.65 (0.52–0.79)
t	0.60 (0.52–0.68)	0.61 (0.52–0.68)	0.60 (0.50–0.71)	0.56 (0.50–0.66)
v	na	na	na	na
ah	0.58 (0.51–0.65)	0.59 (0.52–0.66)	0.55 (0.44–0.66)	0.55 (0.47–0.67)
cg	0.59 (0.32–0.86)	0.54 (0.15–0.76)	na	na
ci	0.61 (0.54–0.67)	0.57 (0.51–0.64)	0.62 (0.51–0.73)	0.60 (0.50–0.82)
ct	0.60 (0.53–0.67)	0.60 (0.53–0.67)	0.61 (0.48–0.73)	0.58 (0.47–0.80)
cv	0.58 (0.51–0.65)	0.59 (0.52–0.66)	0.52 (0.41–0.65)	0.54 (0.42–0.65)
mm	0.53 (0.44–0.59)	0.51 (0.46–0.59)	0.52 (0.45–0.60)	0.52 (0.45–0.60)

AUROC, area under the receiver operator characteristic curve; TPB, transplant protocol biopsy; eGFR, estimated glomerular filtration rate; na: not administered.
*g, early type of allograft glomerulitis; i, interstitial inflammation; t, tubulitis; v, intimal arteritis; ah, arteriolar hyalinosis; cg, allograft glomerulopathy; ci, interstitial fibrosis; ct, tubular atrophy; cv, fibrous intimal proliferation; mm, glomerular mesangial matrix increase.
†AUROC (upper and lower confidence limits).

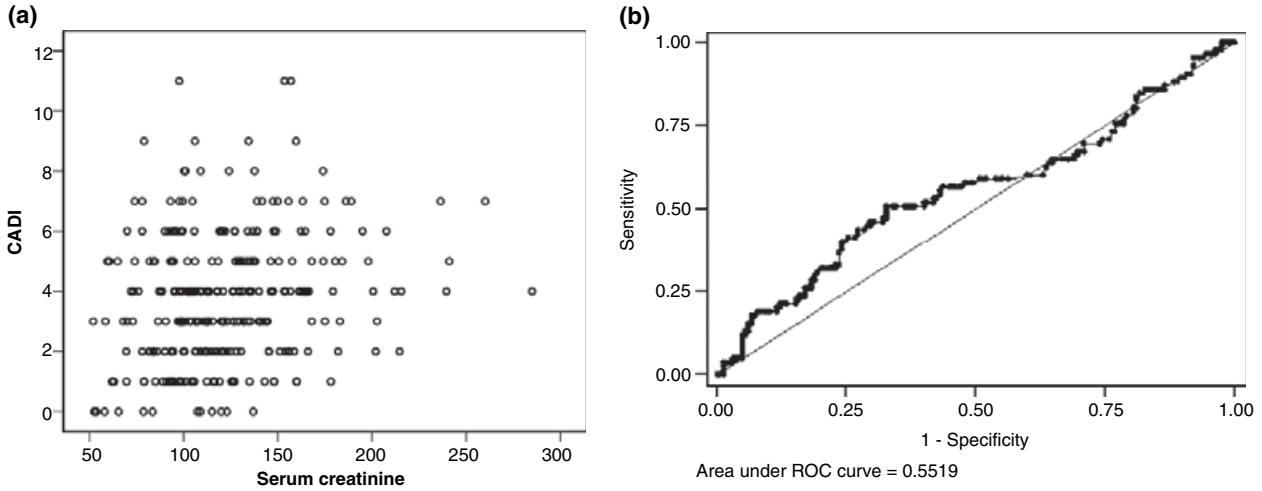

Figure 1 Receiver Operator Characteristic (ROC) curve for serum creatinine as a predictor of chronic allograft damage index (CADI) score >4. Transplant protocol biopsies (TPBs) were taken 6–12 months post-transplant and serum creatinine (independent variable) was measured within 10 days of the biopsy. (a) Scatter plot of serum creatinine versus the TPB CADI score for each patient ($n = 280$), $R^2 = 0.047$. (b) A CADI composite score of >4 for abnormal histopathology was used as the dependent variable. The ROC plot of sensitivity versus 1-specificity (true-positive versus false-positive) for all variable pairs is shown. A test with perfect discrimination (true-positive fraction is 1, false-positive fraction is 0) has an area under the receiver operator characteristic curve (AUROC) of 1, whereas an AUROC of 0.5 (a 45° line) represents a test with no discrimination.

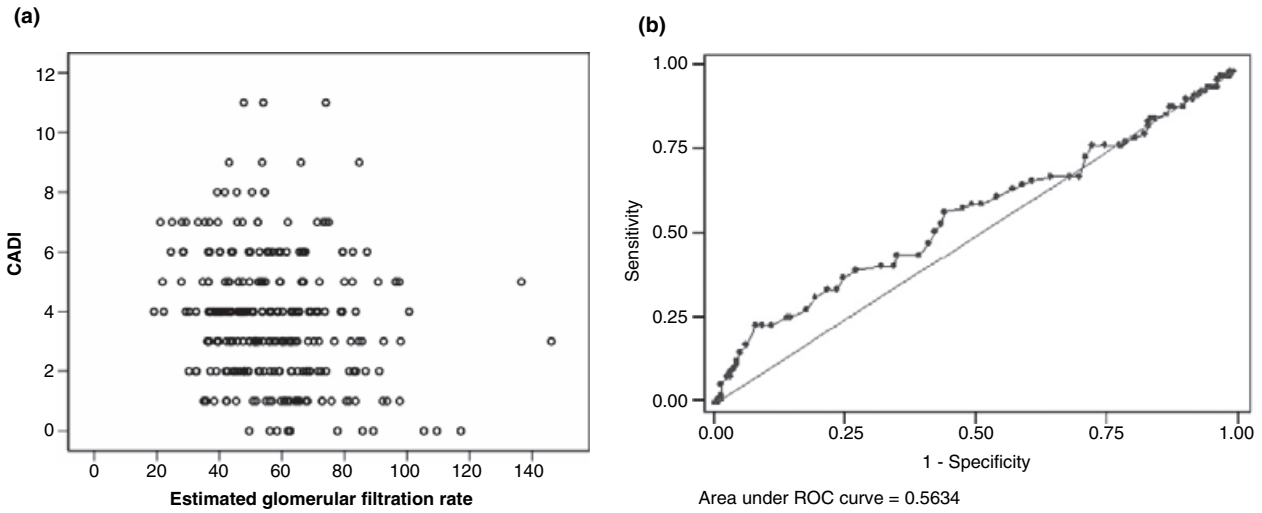

Figure 2 Receiver Operator Characteristic (ROC) curve for estimated glomerular filtration rate (eGFR) as a predictor of chronic allograft damage index (CADI) score >4. Transplant protocol biopsies (TPBs) were taken 6–12 months post-transplant and the eGFR (independent variable) was estimated within 10 days of the biopsy using the Modification of Diet in Renal Disease equation. (a) Scatter plot of eGFR versus the TPB CADI score for each patient ($n = 280$), $R^2 = 0.042$. (b) A CADI composite score of >4 for abnormal histopathology was used as the dependent variable. The ROC plot of sensitivity versus 1-specificity (true-positive versus false-positive) for all variable pairs is shown. A test with perfect discrimination (true-positive fraction is 1, false-positive fraction is 0) has an area under the receiver operator characteristic curve (AUROC) of 1, whereas an AUROC of 0.5 (a 45° line) represents a test with no discrimination.

Discussion

Transport protocol biopsies can provide both sensitive and specific indicators of abnormal renal allograft histopathology, including those changes indicative of early CAN. These histopathological changes both correlate with the clinical end-points of allograft function and survival [6–13,15] and are improved with treatment that also

improves these clinical end-points [14]. Thus, in stable renal transplant recipients, TPBs may help to monitor the safety and effectiveness of novel immunosuppressive regimes as a surrogate marker for CAN [34]. The disadvantage, however, of the use of TPB histopathology is the need to perform an invasive procedure, a concern that has undoubtedly limited the utilization of TPB scoring as a clinical surrogate. As a result, many transplant centres appear to use early functional biomarkers as a replacement for TPB histopathology.

Several retrospective observational studies have demonstrated that serum creatinine [21–23,35] or eGFR [24,35] is correlated with renal allograft survival. However, whereas correlation provides an indication that there is a relationship between two variables, it does not indicate that one variable causes the other. Prediction, however, goes beyond correlation by accounting for some proportion of the variability in the endpoint. Thus, changes in the value of a good predictive marker should, with high sensitivity and specificity, have consequential changes in the endpoint of interest. Serum creatinine and eGFR have not been rigorously evaluated in this regard, nevertheless they have been widely adopted as surrogate markers for CAN in many transplant centres.

The degree to which the putative predictive variable succeeds is best quantified using ROC analysis. This analysis attempts to address the situation where there is a spectrum of possible test results. Reporting a single value for sensitivity and specificity requires selecting a potentially arbitrary cut-off value [28], which may produce an oversimplified and misleading indicator of the accuracy of the test. In contrast, an ROC plot provides a complete picture of a test's accuracy by demonstrating the test's ability to discriminate between alternative health states over the complete spectrum of operating conditions [28,29]. On the ROC plot, all sensitivity versus specificity pairs over the complete range of decision thresholds for the test results are shown by graphing sensitivity (true-positive fraction) and the corresponding value for 1-specificity (false-positive fraction) for each threshold. Although AUROC analysis is an important evaluation tool [29], its limitation is related to the fact that the diagnostic validity of the test is estimated across the whole range of measured values, thus giving equal weight to all false-positive rates [36]. A further limitation of the statistical analysis is biopsy sampling error. Whereas false-positive results due to subcapsular sampling are unlikely when the pathologist is experienced, false-negative results are a risk, due to heterogeneity of histology.

The results of the present study, which used data collected from 280 patients at a single transplant centre and ROC analysis, found that neither serum creatinine nor eGFR could diagnose either individual measures of renal allograft histopathology or a composite measure (CADI score). Consistent with this result, a recent study using ROC analysis demonstrated that serum creatinine at 1 year was a poor predictor of allograft loss at 2 years (AUROC 0.63) [37]. These results illustrate the situation where a test result may have a positive correlation with a clinical end-point or surrogate, but fail as a predictive tool.

Our study is in contrast with Schuck *et al.* [38], who concluded that serum creatinine suggests a Banff CAN grade higher than 1 (AUROC 0.806). Differences between the present study and that of Schuck *et al.* include the sample size (280 vs. 77 TPBs), average serum creatinine at the time of biopsy (122 vs. 202 μmol/l), and, mostly importantly, mean interval time between transplantation and TPB (7.5 vs. 34 months). Our study focuses on changes occurring within the first year following transplant, when histological changes may be occurring but when serum creatinine is almost normal.

The ability of serum creatinine to predict moderate to severe interstitial inflammation in renal allografts (AUROC 0.75) was better than its ability to predict a CADI score of >4 (AUROC 0.55). We had expected to see a higher AUROC value for the predictive value of serum creatinine for interstitial inflammation, because the presence of inflammatory cells, especially macrophages, in renal allografts has been shown to confer a poor prognosis and correlates with fibrosis and vascular sclerosis [39–41].

The ultimate criterion for assessing markers for a pathological condition is whether they add information beyond that otherwise available and whether this information leads to a change in management that is ultimately beneficial to the patient [42]. Despite the obvious attractions of using functional markers as a predictors or surrogates for CAN, if they cannot discriminate between clinically relevant subclasses of subjects (those with normal/abnormal early histopathology or those who will/will not have allograft failure at 2 years), then they have a limited clinical role. In the case of serum creatinine and eGFR, the process of CAN may be quite advanced by the time these functional biomarkers suggest allograft dysfunction, thus limiting the effectiveness of clinical interventions. The risk of TPB as compared with the risk of a possible delay of these interventions needs to be carefully weighed.

There are many cases in the literature where diagnostic tests have been adopted prematurely because they have not been adequately evaluated [43–45]. Both our study (single centre) and that of Kaplan *et al.* [37] (multi-centre) use ROC analysis to conclude that that serum creatinine and eGFR have a limited clinical role in predicting the early histopathological changes that precede CAN and

allograft loss at 2 years respectively. As the transplant community continues to struggle with the challenge of reducing late allograft loss due to CAN and death with a functioning graft, other biomarkers, such as those revealed by studies of genomics and proteomics, will undoubtedly emerge as potential surrogates. It is important that these markers are rigorously evaluated for surrogacy rather than implemented on the grounds of correlation and convenience alone.

Acknowledgements

We thank Dr Ann Kyle for editorial assistance.

Funding sources

Grant support was received from Astellas Inc., Canada.

Authorship

SY: designed study, collected data, analysed data, wrote the paper; II: performed study; MA performed study; MM: collected data; AS: performed study, collected data, wrote the paper; HB: performed study; KM: analysed data, wrote the paper.

References

1. Hariharan S, Johnson CP, Bresnahan BA, Taranto SE, McIntosh MJ, Stablein D. Improved graft survival after renal transplantation in the United States, 1988 to 1996. *N Engl J Med* 2000; **342**: 605.
2. Joosten SA, van Kooten C, Paul LC. Pathogenesis of chronic allograft rejection. *Transpl Int* 2003; **16**: 137.
3. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood)* 2005; **24**: 67.
4. Lachenbruch PA, Rosenberg AS, Bonvini E, Cavaille-Coll MW, Colvin RB. Biomarkers and surrogate endpoints in renal transplantation: present status and considerations for clinical trial design. *Am J Transplant* 2004; **4**: 451.
5. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; **8**: 431.
6. Dimeny E, Wahlberg J, Larsson E, Fellstrom B. Can histopathological findings in early renal allograft biopsies identify patients at risk for chronic vascular rejection? *Clin Transplant* 1995; **9**: 79.
7. Fujisawa M, Ono H, Isotani S, *et al.* Significance of chronic transplant nephropathy on early protocol biopsies for graft outcome in pediatric renal transplantation. *Transplant Proc* 1999; **31**: 1687.
8. Isoniemi HM, Krogerus L, von Willebrand E, Taskinen E, Ahonen J, Hayry P. Histopathological findings in well-functioning, long-term renal allografts. *Kidney Int* 1992; **41**: 155.
9. Kasiske BL, Kalil RS, Lee HS, Rao KV. Histopathologic findings associated with a chronic, progressive decline in renal allograft function. *Kidney Int* 1991; **40**: 514.
10. Nicholson ML, Bailey E, Williams S, Harris KP, Furness PN. Computerized histomorphometric assessment of protocol renal transplant biopsy specimens for surrogate markers of chronic rejection. *Transplantation* 1999; **68**: 236.
11. Nicholson ML, McCulloch TA, Harper SJ, *et al.* Early measurement of interstitial fibrosis predicts long-term renal function and graft survival in renal transplantation. *Br J Surg* 1996; **83**: 1082.
12. Nickerson P, Jeffery J, Gough J, *et al.* Identification of clinical and histopathologic risk factors for diminished renal function 2 years posttransplant. *J Am Soc Nephrol* 1998; **9**: 482.
13. Seron D, Moreso F, Bover J, *et al.* Early protocol renal allograft biopsies and graft outcome. *Kidney Int* 1997; **51**: 310.
14. Mota A, Arias M, Taskinen EI, *et al.* Sirolimus-based therapy following early cyclosporine withdrawal provides significantly improved renal histology and function at 3 years. *Am J Transplant* 2004; **4**: 953.
15. Yilmaz S, Tomlanovich S, Mathew T, *et al.* Protocol core needle biopsy and histologic Chronic Allograft Damage Index (CADI) as surrogate end point for long-term graft survival in multicenter studies. *J Am Soc Nephrol* 2003; **14**: 773.
16. Goldberg BB, Pollack HM, Kellerman E. Ultrasonic localization for renal biopsy. *Radiology* 1975; **115**: 167.
17. Wiseman DA, Hawkins R, Numerow LM, Taub KJ. Percutaneous renal biopsy utilizing real time, ultrasonic guidance and a semiautomated biopsy device. *Kidney Int* 1990; **38**: 347.
18. Furness PN, Philpott CM, Chorbadjian MT, *et al.* Protocol biopsy of the stable renal transplant: a multicenter study of methods and complication rates. *Transplantation* 2003; **76**: 969.
19. Hergesell O, Felten H, Andrassy K, Kuhn K, Ritz E. Safety of ultrasound-guided percutaneous renal biopsy-retrospective analysis of 1090 consecutive cases. *Nephrol Dial Transplant* 1998; **13**: 975.
20. Schwarz A, Gwinner W, Hiss M, Radermacher J, Mengel M, Haller H. Safety and adequacy of renal transplant protocol biopsies. *Am J Transplant* 2005; **5**: 1992.
21. Hariharan S, McBride MA, Cherikh WS, Tolleris CB, Bresnahan BA, Johnson CP. Post-transplant renal function in the first year predicts long-term kidney transplant survival. *Kidney Int* 2002; **62**: 311.
22. Salvadori M, Rosati A, Bock A, *et al.* One-year posttransplant renal function is a strong predictor of long-term kidney function: results from the Neoral-MOST Observational Study. *Transplant Proc* 2003; **35**: 2863.
23. Siddiqi N, McBride MA, Hariharan S. Similar risk profiles for post-transplant renal dysfunction and long-term graft

failure: UNOS/OPTN database analysis. *Kidney Int* 2004; **65**: 1906.
24. Salvadori M, Rosati A, Bock A, *et al*. Estimated one-year glomerular filtration rate is the best predictor of long-term graft function following renal transplant. *Transplantation* 2006; **81**: 202.
25. Katz MH. Common uses of multivariable models. In: Katz MH, ed. *Multivariable Analysis. A Practical Guide for Clinicians*. Cambridge, UK: Cambridge University Press, 1999: 17–26.
26. Sharples LD. Statistical approaches to rational biomarker selection. In: Trull AK, Demers LM, Holt DW, Johnston A, Tredger JM, Price CP, eds. *Biomarkers of Disease: An Evidence-Based Approach*. Cambridge, UK: Cambridge University Press, 2002: 24–31.
27. Lang TA, Secic M. Determining the presence or absence of disease. In: Lang TA, Secic M, eds. *How to Report Statistics in Medicine Annotated Guidlines for Authors, Editors and Reviewers*. Philadelphia, PA: American College of Physicians, 1997: 147–169.
28. Begg CB. Advances in statistical methodology for diagnostic medicine in the 1980's. *Stat Med* 1991; **10**: 1887.
29. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; **39**: 561.
30. Racusen LC, Solez K, Colvin RB, *et al*. The Banff 97 working classification of renal allograft pathology. *Kidney Int* 1999; **55**: 713.
31. Isoniemi H, Taskinen E, Hayry P. Histological chronic allograft damage index accurately predicts chronic renal allograft rejection. *Transplantation* 1994; **58**: 1195.
32. Yilmaz S, Yilmaz A, Hayry P. Chronic renal allograft rejection can be predicted by area under the serum creatinine versus time curve (AUCCr). *Kidney Int* 1995; **48**: 251.
33. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; **20**: 37.
34. Bohmig GA, Regele H, Horl WH. Protocol biopsies after kidney transplantation. *Transpl Int* 2005; **18**: 131.
35. Tesi RJ, Bailly JE, Schnitzler MA, Irish WD. Development of a surrogate marker of long-term kidney allograft outcome based on serum creatinine (sCr) and calculated creatinine clearance at 6 months post-transplant [abstract]. *Am J Transplant* 2002; **2**(**Suppl.3**): 149.
36. McClish DK. Analyzing a portion of the ROC curve. *Med Decis Making* 1989; **9**: 190.
37. Kaplan B, Schold J, Meier-Kriesche HU. Poor predictive value of serum creatinine for renal allograft loss. *Am J Transplant* 2003; **3**: 1560.
38. Schuck O, Viklicky O, Voska L, *et al*. Predicting of histopathological grade of chronic allograft nephropathy from renal function and proteinuria. *Ann Transplant* 2003; **8**: 5.
39. Cosio FG, Grande JP, Wadei H, Larson TS, Griffin MD, Stegall MD. Predicting subsequent decline in kidney allograft function from early surveillance biopsies. *Am J Transplant* 2005; **5**: 2464.
40. Pilmore HL, Painter DM, Bishop GA, McCaughan GW, Eris JM. Early up-regulation of macrophages and myofibroblasts: a new marker for development of chronic renal allograft rejection. *Transplantation* 2000; **69**: 2658.
41. Srinivas TR, Kubilis PS, Croker BP. Macrophage index predicts short-term renal allograft function and graft survival. *Transpl Int* 2004; **17**: 195.
42. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 703.
43. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess* 2005; **9**: 1.
44. Nierenberg AA, Feinstein AR. How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 1988; **259**: 1699.
45. Valenstein PN. Evaluating diagnostic tests with imperfect standards. *Am J Clin Pathol* 1990; **93**: 252.

